# TOPOLOGY-PRESERVING SHAPE-BASED REGRESSION OF RETINAL LAYERS IN OCT IMAGE DATA USING CONVOLUTIONAL NEURAL NETWORKS

*Timo Kepp*[*†]        *Jan Ehrhardt*[*]        *Mattias P. Heinrich*[*]        *Gereon Hüttmann*[‡]        *Heinz Handels*[*]

[*] Institute of Medical Informatics, University of Lübeck, Germany
[†] Graduate School for Computing in Medicine and Life Sciences, University of Lübeck, Germany
[‡] Institute of Biomedical Optics, University of Lübeck, Germany

## ABSTRACT

Optical coherence tomography (OCT) is a non-invasive imaging modality that provides cross-sectional 3D images of biological tissue. Especially in ophthalmology OCT is used for the diagnosis of various eye diseases. Automatic retinal layer segmentation algorithms, which are increasingly based on deep learning techniques, can support diagnostics. However, topology properties, such as the order of retinal layers, are often not considered. In our work, we present an automatic segmentation approach based on shape regression using convolutional neural networks (CNNs). Here, shapes are represented by signed distance maps (SDMs) that assign the distance to the next object contour to each pixel. Thus, spatial regularization is introduced and plausible segmentations can be produced. Our method is evaluated on a public OCT dataset and is compared with two classification-based approaches. The results show that our method has fewer outliers with comparable segmentation performance. In addition, it has an improved topology preservation, which saves further post-processing.

***Index Terms***— Image segmentation, regression, topology, OCT, retina

## 1. INTRODUCTION

Since its introduction in the 1990s, optical coherence tomography (OCT) has become a standard imaging modality for retinal diagnostics in ophthalmology. OCT enables the non-invasive acquisition of image volumes of the retina in micrometer-resolution. Due to the high level of detail in the depictive representation by OCT, it is particularly suitable to detect pathological structures within the retina, which can be used as biomarkers e.g. the individual retinal layer thickness or the volume of diseased fluid deposits (see Figure 1). The presence of biomarkers and their change over time is associated with certain retinal diseases, such as age-related macular degeneration (AMD) or diabetic macular edema (DME). The procedure of measuring biomarkers in OCT image data requires manual segmentation of the respective structures in the OCT image. Manual segmentation of biomarkers in OCT
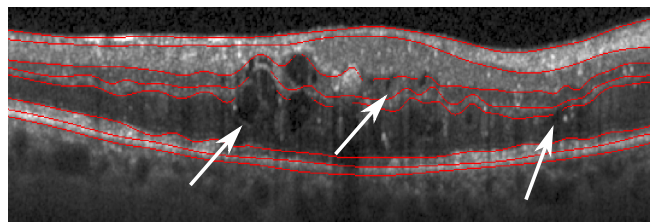


**Fig. 1**: OCT scan of a patient's retina with diabetic macular edema (DME). The disease causes thickness changes of the individual retinal layers (red annotations). In addition, fluid deposits (white arrows) leads to further disorders.

images is very time consuming and therefore not appropriate in clinical practice. In order to save time as well as personnel resources, the development of automatic segmentation methods has been motivated. In recent years, algorithms for the segmentation of retinal layers and fluid deposits within the retina have been presented in several works. These algorithms often use machine learning [1] and graph theoretical models [2, 3]. Although graph theoretical models guarantee a correct topology of the retinal layers, they are very complex to create and have to be adapted for different disease types.

More recent algorithms are based on deep learning methods, convolutional neural networks (CNNs) in particular. Based on their training data CNNs are able to automatically learn representative features in order to built very complex classification or regression models. A very popular CNN architecture in medical image analysis is the U-Net [4]. In their work Roy et al. [5] presented a U-Net variant and trained it with a joint loss function in order to segment both retinal layers and fluid deposits. He et al. [6] used a cascade of two U-Nets to segment retinal layers. Here, the first network segments the retina, whereas the second one corrects segmentation errors. Due to the fact that both methods perform a pixel-wise classification, topology, e.g. in form of shapes, can not be considered.

Further works integrated shape and topology constraints in their segmentation algorithms, respectively. Fang et al. [7] segmented nine retinal layers in OCT image data from pa-

tients with non-exudative AMD. They linked a CNN model to a graph search method [2] in order to provide topology constraints. Furthermore, He et al. [8] proposed an advanced version of their previous approach [6]. Again, the first network performs a pixel-wise classification. However, the second network conducts a regression of the individual retinal layer thicknesses, by which topologically correct retinal segmentation can be guaranteed. Even tough topology properties are considered, the employed network architectures are complex and used in combination with additional mathematical models [7], or require further computational effort [8].

In this work, we present a new approach for the segmentation of retinal layers and fluid in OCT image data based on a regression-based segmentation approach using CNNs. Instead of constructing complex network architectures or loss functions, we regress shape representations of the retinal layers as well as fluid deposits. This ensures spatial regularization creating plausible segmentations. Glocker et al. [9] introduced a combination of pixel-wise classification and shape regression using decision forests. The shape information was implicitly represented by signed distance maps (SDMs). The segmentation results showed increased consistency and robustness. SDMs assign the distance to the object contour to each pixel and indicate whether the pixel is inside or outside the object. They can easily be generated from the existing annotated labels. Heinrich et al. [10] recently achieved significantly improved segmentation results by performing a CNN-based regression of SDMs and reached the first place in the CREMI challenge for synaptic cleft segmentation. Due to the advantages described above, we also employ SDMs for our regression task. For this purpose, the approach in [10] is extended to a multi-class regression in order to segment seven retinal layers as well as fluid deposits in OCT image data of DME patients.

## 2. MATERIALS AND METHODS

**Shape-based regression:** The main contribution of this work is a shape-based regression of retinal layers and fluid deposits in OCT image data of DME patients using CNNs. In contrast to classification-based segmentation with each pixel assigned to a predefined category, we deploy a regression-based approach. To ensure spatial consistency we employ SDMs, which can easily be computed from ground truth label data. Furthermore, regression of SDMs ensures spatial regularization and thus preserves topology. SDMs are defined as follows:

$$SDM(\boldsymbol{x}) = \begin{cases} -\min_{\boldsymbol{y} \in \Omega_B} d(\boldsymbol{x}, \boldsymbol{y}), & \text{if } \boldsymbol{x} \in \Omega_F \\ \min_{\boldsymbol{y} \in \Omega_F} d(\boldsymbol{x}, \boldsymbol{y}), & \text{if } \boldsymbol{x} \in \Omega_B. \end{cases} \quad (1)$$

$\Omega_F$ and $\Omega_B$ denote the set of pixels belonging to the foreground (inside the segmentation mask) and background respectively. $d(\boldsymbol{x}, \boldsymbol{y})$ represents the distance metric. In this work, the Euclidean distance $d(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{\|\boldsymbol{x} - \boldsymbol{y}\|^2}$ is used. As suggested in [10], we scale the SDM and apply a $\tanh$ non-linearity: $SSDM = \tanh(\lambda SDM)$. This takes areas with large distances to object contours less into account and thus simplifies the regression task. We empirically have determined a value of $0.1$ for $\lambda$. In order to obtain the respective segmentation labels, we compute the minimum for each pixel over all SSDM channels. No further post-processing is utilized.

**Network architecture:** In order to evaluate the influence of shape-based regression independently of the network architecture, we use the widely used U-Net architecture [4]. Like every fully convolutional network (FCN), the U-Net consists of a contracting encoder and a expanding decoder part. The encoder gradually reduces the spatial resolution using strided max pooling, whereas the decoder recovers object details as well as spatial resolution. The encoder detects spatial context in different resolution levels, which is combined with the features of the decoder via skip connections. This allows the decoder to recover relevant features lost by the encoder's pooling operations.

We have replaced the transposed convolutions with bilinear upsampling operations in order to avoid checkerboard artifacts. In doing so the amount of trainable parameters was reduced. In addition, the upsampling operation is followed by a $1 \times 1$ convolution to receive the same number of feature channels compared to the transposed convolution. The number of feature channels at the highest scale level is set to $64$ which results in $1024$ channels at the lowest level. The network receives as a input a single-channel image and outputs an image, whose number of channels corresponds to the number of classes.

## 3. EXPERIMENTS AND RESULTS

**Dataset:** The presented approach is evaluated on the publicly available Duke OCT dataset [11], which is also used in [5]. The dataset consists of ten OCT scans from different patients with DME. Two medical experts manually annotated retinal layers and fluid deposits of the 11 centered slices (B-scans) of each OCT image. This results in 110 B-scans per expert. The following nine structures have been segmented: inner limiting membrane (ILM ■), nerve fiber layer to inner plexiform layer (NFL-IPL ■), inner nuclear layer (INL ■), outer plexiform layer (OPL ■), outer nuclear layer to inner segment myeloid (ONL-ISM ■), inner segment ellipsoid (ISE ■), outer segment to retinal pigment epithelium (OS-RPE □), fluid (■) and background. The image volumes were acquired by a Spectralis OCT scanner (Heidelberg Engineering GmbH) where each B-scan has a size of $512 \times 740$ pixels. Further details regarding this dataset are reported in [11].

Due to the large OCT B-scan size, we halved the image size for computational efficiency and performed zero padding, which results in a resolution of $256 \times 288$ pixels for

**Table 1**: Averaged metric results of the leave-one-out cross validations. Best performance is shown in bold. Values in columns $E_1$ and $E_2$ represent metric results that have been computed using the annotated ground truth of the first and the second expert, respectively. In addition, the inter-rater reliability (IRR) between both experts has been determined.

| Anatomic labels → | | ILM ■ | | NFL-IPL ■ | | INL ■ | | OPL ■ | | ONL-ISM ■ | | ISE ■ | | OS-RPE □ | | Fluid ■ | | Average | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Metrics↓ | Methods↓ | $E_1$ | $E_2$ | $E_1$ | $E_2$ | $E_1$ | $E_2$ | $E_1$ | $E_2$ | $E_1$ | $E_2$ | $E_1$ | $E_2$ | $E_1$ | $E_2$ | $E_1$ | $E_2$ | $E_1$ | $E_2$ |
| DC | U-Net-L1 | 0.81 | 0.79 | 0.85 | 0.84 | 0.72 | 0.72 | 0.71 | 0.66 | 0.87 | 0.85 | 0.84 | 0.82 | 0.83 | 0.81 | **0.46** | 0.40 | 0.76 | 0.74 |
| | U-Net-DC | **0.84** | 0.80 | **0.88** | 0.85 | 0.74 | **0.76** | 0.73 | **0.70** | **0.88** | 0.85 | 0.84 | **0.83** | **0.84** | **0.82** | 0.39 | 0.36 | 0.77 | 0.75 |
| | U-Net-CE | **0.84** | **0.82** | **0.88** | **0.86** | **0.77** | 0.75 | **0.74** | 0.68 | 0.88 | **0.87** | **0.85** | **0.83** | **0.84** | 0.81 | 0.42 | **0.45** | **0.78** | **0.76** |
| | IRR | 0.85 | | 0.89 | | 0.77 | | 0.72 | | 0.87 | | 0.85 | | 0.84 | | 0.49 | | 0.79 | |
| ASSD | U-Net-L1 | 1.03 | 1.83 | 1.25 | 1.27 | 1.21 | 1.51 | **1.24** | 1.62 | 1.41 | 1.71 | **0.60** | 0.68 | **0.59** | **0.68** | – | – | 1.04 | 1.33 |
| | U-Net-DC | **0.81** | 1.69 | 1.06 | **1.13** | 3.17 | **1.16** | 1.86 | 1.95 | 1.34 | 1.54 | 0.62 | **0.66** | 0.62 | 0.70 | – | – | 1.35 | 1.26 |
| | U-Net-CE | 0.83 | **1.35** | **0.99** | 1.42 | **1.12** | 1.59 | 1.34 | **1.42** | **1.26** | **1.42** | 0.61 | **0.66** | **0.59** | 0.69 | – | – | **0.96** | **1.22** |
| | IRR | 1.70 | | 1.76 | | 1.91 | | 1.87 | | 2.17 | | 1.17 | | 1.19 | | – | | 1.68 | |
| HD | U-Net-L1 | 9.56 | 11.74 | **7.45** | **6.77** | 8.50 | 9.62 | 9.59 | 11.26 | 10.78 | 12.10 | **2.79** | **2.87** | 2.59 | 2.87 | – | – | 7.36 | 8.16 |
| | U-Net-DC | 9.87 | 16.87 | 8.39 | 9.53 | 16.53 | 8.69 | 12.63 | **10.31** | 11.35 | 11.84 | 5.76 | 4.86 | 6.03 | 4.92 | – | – | 10.08 | 9.57 |
| | U-Net-CE | **7.11** | **11.46** | 8.86 | 7.99 | 9.04 | 9.67 | 10.32 | 11.03 | **9.84** | **11.69** | 4.16 | 3.61 | 3.30 | 3.49 | – | – | 7.52 | 8.42 |
| | IRR | 13.60 | | 9.33 | | 11.78 | | 11.18 | | 14.78 | | 4.48 | | 4.38 | | – | | 9.93 | |
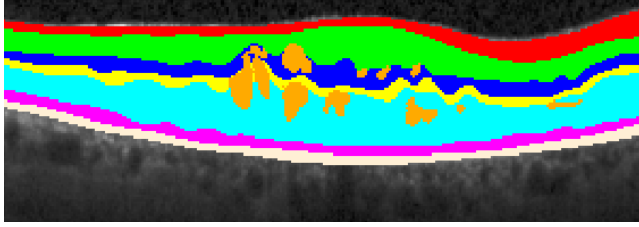


(a) Ground truth annotation by expert 1
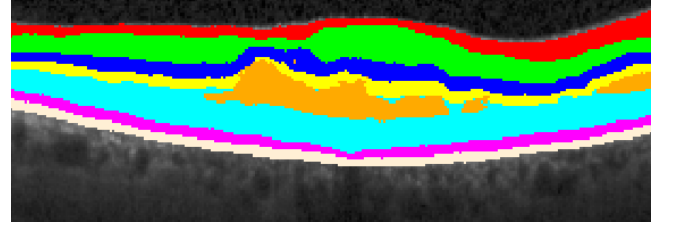


(b) U-Net-L1 predictions (proposed)
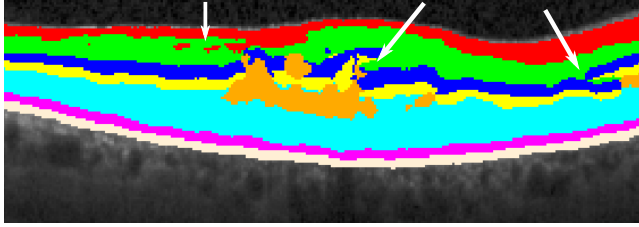


(c) U-Net-CE predictions
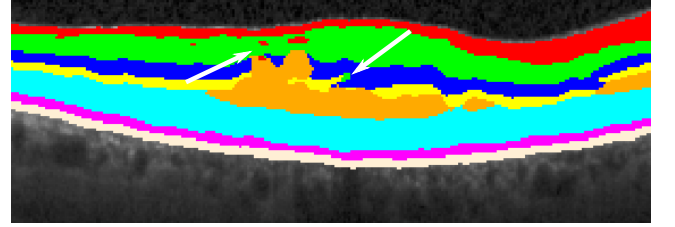


(d) U-Net-DC predictions

**Fig. 2**: Test OCT B-scan overlaid with ground truth annotations (expert 1) (a), with predictions of U-Net-L1 (b) and with predictions of U-Net-CE (c) and U-Net-DC (d), respectively. Color mapping is explained as follows: ILM ■, NFL-IPL ■, INL ■, OPL ■, ONL-ISM ■, ISE ■, OS-RPE □ and fluid ■. Topology errors are indicated by white arrows. The same B-scan as in Figure 1 is shown.

each B-scan.

**Training:** Adam optimization is used to minimize a L1 loss with respect to the SSDMs of the binarized labels. The initial learning rate is set to $0.001$. In addition, an exponential scheduler is utilized to adjust the learning rate during the training where a decay of $\gamma = 0.99$ is used. Furthermore, we apply $\beta_1 = 0.9$ and $\beta_2 = 0.999$ as exponential decay rates for the moment estimates. Moreover, a mini-batch size of five B-scans is used to optimize the throughput of the GPU. The network model is trained for 200 epochs on a NVidia GeForce GTX 1080 TI graphics card using the PyTorch framework.

Because of the small number of training samples, we carry out an online data augmentation. In addition to simple horizontal flips and rotations ($\pm 10°$), we also apply elastic transformations to make the network less susceptible to atypical

morphologies e.g. in regions with fluid deposits.

After each epoch we compute the average symmetric surface distance (ASSD) between the contours of the predicted SSDMs and the ground truth labels. The model that achieves the lowest ASSD for all class labels on the validation set at a particular epoch is used for testing.

**Evaluation:** We test our regression-based segmentation (**U-Net-L1**) against two classification-based approaches using cross entropy (**U-Net-CE**) and dice coefficient (**U-Net-DC**) as loss function, respectively. A leave-one-out cross-validation is employed to evaluate our approach. For this purpose, a network is trained per patient whose B-scans are used as a test set (one subject, 11 B-scans) in each case. The remaining patient datasets are divided into training (seven subjects, 77 B-scans) and validation sets (two subjects, 22

B-scans). Since two experts each annotated all OCT B-scans, the leave-one-out cross-validation was performed for both ground truth datasets.

In order to measure the segmentation accuracy, we use three different metrics: Dice coefficient (DC), average symmetric surface distance (ASSD) and Hausdorff distance (HD). The latter is used to detect outliers. The evaluation results are shown in Table 1. Concerning DC and ASSD, our proposed method achieves a slightly lower performance in comparison to **U-Net-CE** and **U-Net-DC**. Looking at the HD values, one can see that **U-Net-L1** has on average the smallest distances. Furthermore, it should be emphasized that **U-Net-L1** segments fluid deposits with comparable accuracy to **U-Net-CE/DC** (c.f. Table 1). An example segmentation of the different methods is shown in Figure 2. It demonstrates that **U-Net-CE/DC** (Figures 2c and 2d) shows segmentation errors with regard to the topology. Corresponding areas are marked with arrows (c.f. 2c/2d). In contrast, our proposed method **U-Net-L1** (Figure 2b) shows no topology errors. To further demonstrate the topology preservation of our method, we applied a connected component filter to each retinal layer class. The filter should output only a single component for each retinal layer class unless it is separated by fluid. The percentage deviation of the computed components with respect to the ground truth has been determined. Here, our method achieves on average a lower deviation (0.26) from the ground truth than **U-Net-CE/DC** (0.29/0.33). Although **U-Net-L1** preserves topology, individual retinal layers appear slightly smoothed, which explains the higher ASSDs.

## 4. CONCLUSION

In this work, we have presented a new approach for the segmentation of retinal layers and fluid deposits in OCT image data. We have shown that regression of SSDMs leads to robust segmentation results and to improved topology preservation at the same time.

Compared to other works [7, 8] our method does not require additional mathematical models or refinement steps to produce topologically correct segmentations. In addition, our approach also segments fluid deposits with a comparable accuracy with respect to other methods (**U-Net-CE/DC**). Despite the slightly lower performance regarding DC values and ASSDs, our proposed method produces robust and topologically correct segmentations, c.f. HD, which are clinically more important than local pixel-wise accuracy. In future work we are interested to improve the performance of our method with respect to DC and ASSD, e.g. by extending the loss function. Moreover, our approach will be evaluated using larger image datasets.

## 5. REFERENCES

[1] A. Lang, A. Carass, et al., "Retinal layer segmentation of macular OCT images using boundary classification," *Biomed. Opt. Express*, vol. 4, no. 7, pp. 1133–1152, 2013.

[2] S. J Chiu, X. T. Li, et al., "Automatic segmentation of seven retinal layers in SDOCT images congruent with expert manual segmentation," *Opt. Express*, vol. 18, no. 18, pp. 19413–19428, 2010.

[3] R. Kafieh, H. Rabbani, et al., "Intra-retinal layer segmentation of 3D optical coherence tomography using coarse grained diffusion map," *Medical Image Analysis*, vol. 17, no. 8, pp. 907–928, 2013.

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.

[5] A. G. Roy, S. Conjeti, et al., "ReLayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks," *Biomed. Opt. Express*, vol. 8, no. 8, pp. 3627–3642, 2017.

[6] Y. He, A. Carass, et al., "Towards Topological Correct Segmentation of Macular OCT from Cascaded FCNs," in *Fetal, Infant and Ophthalmic Medical Image Analysis*, pp. 202–209. Springer, 2017.

[7] L. Fang, D. Cunefare, et al., "Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search," *Biomed. Opt. Express*, vol. 8, no. 5, pp. 2732–2744, 2017.

[8] Y. He, A. Carass, et al., "Topology guaranteed segmentation of the human retina from OCT using convolutional neural networks," *arXiv preprint arXiv:1803.05120*, 2018.

[9] B. Glocker, O. Pauly, et al., "Joint classification-regression forests for spatially structured multi-object segmentation," in *European Conference on Computer Vision*. Springer, 2012, pp. 870–881.

[10] L. Heinrich, J. Funke, et al., "Synaptic Cleft Segmentation in Non-Isotropic Volume Electorn Microscopy of the Complete Drosophilia Brain," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 317–325.

[11] S. J. Chiu, M. J. Allingham, et al., "Kernel regression based segmentation of optical coherence tomography images with diabetic macular edema," *Biomed. Opt. Express*, vol. 6, no. 4, pp. 1172–1194, 2015.